



OPEN ACCESS

EDITED BY

Naglaa Ali Megahed,
Port Said University, Egypt

REVIEWED BY

Asmaa M. Hassan,
Independent Researcher, Port Said, Egypt
Metin Bulus,
Akiyaman University, Türkiye

*CORRESPONDENCE

Anne G. M. de Bruijn
✉ a.g.m.de.bruijn@vu.nl

RECEIVED 21 September 2023

ACCEPTED 27 November 2023

PUBLISHED 19 December 2023

CITATION

de Bruijn AGM and Meeter M (2023) Catching up after COVID-19: do school programs for remediating pandemic-related learning loss work?

Front. Educ. 8:1298171.

doi: 10.3389/feduc.2023.1298171

COPYRIGHT

© 2023 de Bruijn and Meeter. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Catching up after COVID-19: do school programs for remediating pandemic-related learning loss work?

Anne G. M. de Bruijn* and Martijn Meeter

LEARN! Research Institute, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Introduction: COVID-19 had a major impact on education, resulting in learning losses among students. The Dutch ministry set-up a subsidy for schools to implement catch-up programs in tackling learning losses. In this study, we examine (a) which students participated in the programs, and (b) effectiveness of these programs in remediating learning losses in secondary school students.

Methods: Sixteen program in eight secondary schools were analyzed using data of 16,675 students (9,784 individual students; 1,336 participating in a catch-up program). Schools implemented three program types: tutoring, homework support, and general skills. Per school, a difference-in-difference design was used, computing two effect sizes: comparing grades of participating and non-participating students; and grades in tutoring-specific subjects to non-tutored subject (specifically for tutoring programs). Effect sizes were combined using meta-analytic regressions in JASP.

Results: At program onset, students selected for participation had significantly lower overall grades than non-participants, or – for subject-specific tutoring – lower grades specifically in the tutored subject. Tutoring programs significantly increased students' grades: with higher grades for participants than non-participants, and – for students receiving subject-specific tutoring - higher grades in tutored subjects compared to those in non-tutored subjects. No significant effects were found for homework support and general study skill programs.

Conclusion: Schools selected students most in need for catch-up programs. Tutoring interventions seemed to remediate part of secondary school students' learning losses, whereas general skills programs and homework support programs did not. Large between-school heterogeneity was found, implying that program implementation was at least as important as program type and content.

KEYWORDS

learning losses, COVID-19, program effectiveness, secondary education, catch-up program, tutoring

Introduction

Worldwide, lockdowns due to COVID-19 have had an enormous impact on education, with school closures resulting in large learning losses among primary and secondary school students (Hammerstein et al., 2021; Zierer, 2021; König and Frey, 2022; Betthäuser et al., 2023; Di Pietro, 2023). Results of meta-analyses indicate that students lost around 18–35% of a normal school

year worth of learning (Zierer, 2021; König and Frey, 2022; Betthäuser et al., 2023). Also in the Netherlands, learning gains during the period of school closures were found to be lower than they would otherwise have been (Engzell et al., 2021; Haelermans et al., 2022a; Schuurman et al., 2023). Worryingly, learning losses seemed to have been largest for students from disadvantaged backgrounds (Engzell et al., 2021; Hammerstein et al., 2021; Haelermans et al., 2022a; Betthäuser et al., 2023; Schuurman et al., 2023). Although learning deficits have not widened substantially since the early pandemic, nor has inequality among students; children and adolescents still seem to be facing adverse effects of school closures (Haelermans et al., 2022b; Betthäuser et al., 2023). Also, learning delays are still largest among the most disadvantaged children (Haelermans et al., 2022c; Betthäuser et al., 2023).

Subsidy arrangement in the Netherlands

In the Netherlands, the Ministry of Education, Culture and Science decided to set up a subsidy arrangement to help schools in remediating the experienced learning losses. In the period after the first lockdown, schools could apply for subsidy to implement catch-up programs for their most vulnerable students. Multiple application rounds were provided, and schools could implement multiple catch-up programs simultaneously. Unique about the Dutch program were the speed with which it was set up (first subsidies paid out in September 2020) and the freedom schools had in designing the programs. They were free in determining what type of program they implemented, what the goals of this program were, who implemented the program, and which students they included for participation (as long as these were the students most in need of support as a result of the lockdowns). The resulting variance in programs provides a unique opportunity to study the effectivity of different program types in combating learning losses due to lockdowns.

Most Dutch primary schools (around 70%) and secondary schools (around 90%) applied for the subsidy. In line with the emphasis often being laid on academic goals, most schools focused on fostering students' academic achievement in the core subjects (mathematics and Dutch/English language; De Bruijn et al., 2021). In reaching these goals, schools often made use of tutoring or extension of school hours, followed by homework support or general skills training (e.g., exam training; De Bruijn et al., 2021). Other types of interventions, such as social-emotional support or learning-to-learn training were also tried. More information on the program choices schools made can be found in our previous publications (Kortekaas-Rijlaarsdam et al., 2020; De Bruijn et al., 2021).

Related work

Results on the effectiveness of the catch-up programs in primary schools indicated that they reduced, although not completely remediated, students' learning losses (Haelermans et al., 2021). That is: differences in learning growth between participants in the catch-up programs and those not participating became smaller, yet did not disappear. Although programs were also widely implemented in secondary schools, the effectiveness of these programs in reducing learning losses caused by COVID-19 have not yet been examined,

neither in the Netherlands, nor internationally. Also, it remains unknown whether program effectiveness differed depending on the type of program being implemented. Meta-analytic studies have shown that, in general, tutoring interventions are among the most effective in enhancing learning, reporting medium to large (Kraft, 2020) effect sizes (ES=0.24; Baye et al., 2019; ES=0.36; Dietrichson et al., 2017; ES=0.26; Inns et al., 2019; ES=0.37; Nickow et al., 2020; 0.20 – Pellegrini et al., 2021), and supporting schools' preference for implementing tutoring programs to remediate learning losses. Still, most studies examining effectiveness of academic interventions have been conducted among primary school students (Inns et al., 2019; Nickow et al., 2020; Pellegrini et al., 2021). Also, studies have not yet examined effectiveness of this type of program when specifically aiming to combat learning losses caused by COVID-19. In light of the adverse effects of COVID-19-related school closings that are still existing amongst students, it seems of vital importance to gain insight into the effectiveness of catching-up programs in combating learning losses; and the program type that is the most effective in doing so. This way, inequality among students can be reduced, and students can be helped in making the most of their academic career.

The present study

Therefore, the main aim of this study is to examine the effectiveness of catch-up programs in combating COVID-19-related learning losses of Dutch secondary school students. In doing so, we have three specific research questions:

- (1) Did schools select those students for catch-up programs that were most in need of it?
- (2) Were the programs effective in improving learning outcomes among participants?
- (3) Did program effectiveness vary as a function of the type of program?

Based on results of studies on the effectiveness of similar catch-up programs among Dutch primary school students, and meta-analyses on the effectiveness of general academic interventions, we expect that programs will be effective in remediating learning losses, with the strongest effects for tutoring programs.

Materials and methods

Context and sampling

Dutch secondary education starts at grade 7, when students are on average 12 years old. It is highly tracked with students divided in seven tracks, from pre-vocational to pre-university, that differ in duration (from 4 to 6 years) and in the type of higher education they give access to. Secondary schools are quite large, with on average 1700 students, and mostly offer multiple tracks. Schools have large autonomy in both the content and manner in which teaching takes place, with only having central exams in the last secondary school year (OECD, 2018). The subsidy arrangement was designed with this autonomy in mind. Schools had the freedom to design their programs as they saw fit, with the restriction that these had to be targeted at the

10% most affected students (although spillover to other students was allowed, e.g., when learning materials were bought). Schools with many disadvantaged students were allowed to target 20% of their students. Subsidies were available from July 2020, and programs could run up to December 2021.

One of the preconditions for obtaining the subsidy was that schools were obliged to participate, when requested, in evaluation research. For this evaluation, schools were randomly sampled from the list of schools that had requested the subsidy using a random generator in Excel, with the restriction that only schools with at least 900 students and at least 90 confirmed program participants could participate. Twenty-two schools were initially contacted via e-mail, of which 13 agreed to cooperate. Of these, 8 turned out to have programs that were amenable to quantitative analysis of effects. Other schools either had programs that targeted all students, making comparison of participating and non-participating students impossible, did not register participants, or did not use grading. All students at participating schools were included in the study.

Participants

In total, 8 Dutch secondary schools participated. All were large schools providing education in various academic tracks, being located in different parts of the Netherlands. To determine effectiveness of the programs, data of 16,675 students was analyzed (9,784 individual students, taking into account students that were in multiple samples when schools implemented more than one program). Schools were asked to provide anonymized data at student level on gender, grade, grade level, attendance of catch-up program, and school grades. They did so at the start and end of the intervention program. To guarantee anonymity, no further data at student level was requested. Of the analyzed students, 1,336 (8.0%) took part in a catch-up program (note that students may have participated in multiple programs). The project was approved by the Ethical Board of the Faculty of Behavioral and Human Movement Sciences of the Vrije Universiteit Amsterdam (approval number VCWE-2019-151).

Design

For all schools, a difference-in-difference design was used, comparing the improvement in grades of participating students and their non-participating peers. For interventions targeting specific school subjects, we additionally compared the development of grades for program-specific subjects to grades for other subjects. Because schools differed substantially in the way interventions and subjects were chosen, programs were organized, students were selected, and grades were registered, analyses were designed and ran per school and then combined using meta-regression. In total, 16 catch-up programs implemented in the 8 participating secondary schools were analyzed.

Materials

Academic achievement

Students' grades before and after the program – provided by schools – were used as indicator of academic achievement. Although

Dutch secondary schools are free in how they assess learning, most tend to administer summative tests every few weeks or months in all subjects, using the average of obtained grades as report grades. We used as many measurements of grades as possible, taking into account the constraints of the school's grade registration. For most schools, this implied that just two report grades were available: one for the period preceding the intervention and one for the period during or after the intervention. Depending on the catch-up program's goal and content, we examined an overall average of students' grades, or the average grade in a specific academic subject (i.e., that was the target of the catch-up program in which the student was participating).

Catch-up programs

To categorize the catch-up programs implemented by schools, data was used on the type of catch-up program implemented, using information from interviews with school leaders and staff involved in implementing the program (De Bruijn et al., 2021). Programs were categorized into three broader categories: (1) 'tutoring', in which students received individual or small-group instruction in one particular subject; (2) homework support, where students were offered the opportunity to do their homework at school with staff being around for help; and (3) general skills, which consisted of courses in either 'learning to learn', stress reduction, or reading comprehension skills. For schools that implemented multiple programs simultaneously, each individual program was categorized – resulting in multiple program types for one school.

Analyses

For our first aim, examining whether mainly disadvantaged students participated in the programs, we compared grades of participants in the catch-up programs at the onset of the program to those of their non-participating peers. More specifically, we looked whether there were any non-participating students who had lower grades than the average grade of students that were participating in the catch-up programs. For each program, we calculated the percentage of non-participating students with lower grades than the average grade of students participating in the program. A higher percentage indicates that the program was less successful in reaching students in more need of the program (i.e., with lower achievement). For each program we present these percentages in the results section, with 50% indicating that there was no significant difference in grades between participating and non-participating students (i.e., student grade did not seem to be an accurate selection criterion for participation in the program).

Secondly, we examined effects of the catch-up programs on academic achievement. These effects were examined for each of the 16 programs separately, calculating a mixed interaction between group (program participants vs. non-participants) and time (pre- and post-intervention). For schools that implemented multiple programs simultaneously, multiple effect sizes were computed (one for each program). Two effect sizes (Hedges' g) were used: (1) quantifying the difference in learning growth between participants in the catch-up programs and their non-participating peers; (2) specifically for programs that targeted a specific school subject, quantifying the difference in learning growth for the targeted school subject compared to learning growth in other subjects. Here, learning growth refers to

changes in student grades during the intervention, i.e., a comparison of grades before and after the intervention program. Although we initially aimed to examine interactions with grade, track, and initial (pre-intervention) school grade as well, the effects of school subject turned out to be too strong – resulting in estimates of interaction effects that were impossible to interpret (i.e., they varied largely depending on school subject). Therefore, we only took into account differences over school subjects. For interventions that targeted multiple school subjects, the effect of the intervention was computed separately per school subject, and then averaged by weighting the number of participants in the intervention program, this way taking into account different numbers of participants depending on school subject.

Following, meta-analytic regressions were performed in JASP (JASP Team, 2023), one for the comparison between participants and nonparticipants and one for the within-student comparison between subjects with and without special help, averaging effects over schools.

Results

Descriptives of catch-up programs

Table 1 presents an overview of the type of programs implemented at individual schools and the number of students participating. All schools implemented subject-specific tutoring ($n = 10$ programs) often focused on English, mathematics, economics, or (one of the) physical sciences. Two schools used programs aimed at general skills, such as study skills or reading comprehension ($n = 4$ programs); and two schools used programs focused on homework support ($n = 2$ programs).

Participating students

Firstly, we examined whether, at the onset of the program, academic performance of participants in the programs differed significantly from that of their non-participating peers, by analyzing the distribution of student grades (i.e., the percentage of students not participating in a program with lower grades than the average of the participating students). Figure 1 presents these percentages for each of the 16 programs, with lighter circles representing the percentage for average grades over all subjects, and darker circles the percentage for grades in tutoring-specific subjects (note: some schools implemented subject-specific tutoring for multiple subjects, meaning that the total number of circles differs from the total number of programs that were examined).

Overall, grades of students participating in the catch-up programs were lower than those of their non-participating peers. Schools thus seemed to have selected students with larger learning delays to participate in the catch-up programs. For programs aimed at a specific subject, there were few non-participating students with lower grades for that tutoring-specific subject than the average grade of participating students, see the darker circles in Figure 1. However, in half of the schools that had implemented tutoring, the average grade of participating students over all other subjects was not significantly lower than that of non-participating students, see the lighter circles in Figure 1. Thus, for subject-specific programs, schools seem to have

TABLE 1 Overview of the 16 programs implemented in the eight participating schools, the total number of students per school, and the number and percentage of program participants.

School	Program	Students (total)	Program participants (%)
1	Tutoring	839	29 (3.5)
2	Tutoring	1,104	320 (29.0)
3	Tutoring English language	826	20 (2.4)
	Tutoring Mathematics		14 (1.7)
4	Tutoring	276	73 (26.4)
5	Tutoring	203	56 (27.6)
	Study skills – general		26 (12.8)
6	Homework support	1,402	16 (1.1)
	Study skills – reading comprehension		10 (0.7)
	Study skills – stress reduction		16 (1.1)
	Study skills – general		69 (4.9)
	Tutoring		241 (17.2)
7	Homework support	254	48 (18.9)
	Tutoring English language		46 (18.1)
8	Tutoring – lower grades	2,748	112 (4.1)
	Tutoring – higher grades		240 (11.3)
	Total in analysis	16,675	1,336 (8.0)

selected students who were behind in the specific goal-domain, rather than students who were struggling to keep up academically more generally.

Program effects

Secondly, we examined the effects of the three types of catch-up programs by combining the effect sizes for individual programs in a meta-analysis. The first program type, tutoring, was successful in reaching its targeted outcomes. Participants of tutoring programs reached significantly higher grades compared to non-participants (Hedges' $g = 0.21$, $p = 0.013$; see Figure 2). Also, positive effects were found when comparing students' grades in the specific subject that was the focus of the program to students' grades in all other subjects (overall estimate of Hedges' $g = 0.26$, $p = 0.038$; see Figure 3). Students participating in a subject-specific tutoring program improved their grades in the specific subject they were tutored in, compared to their

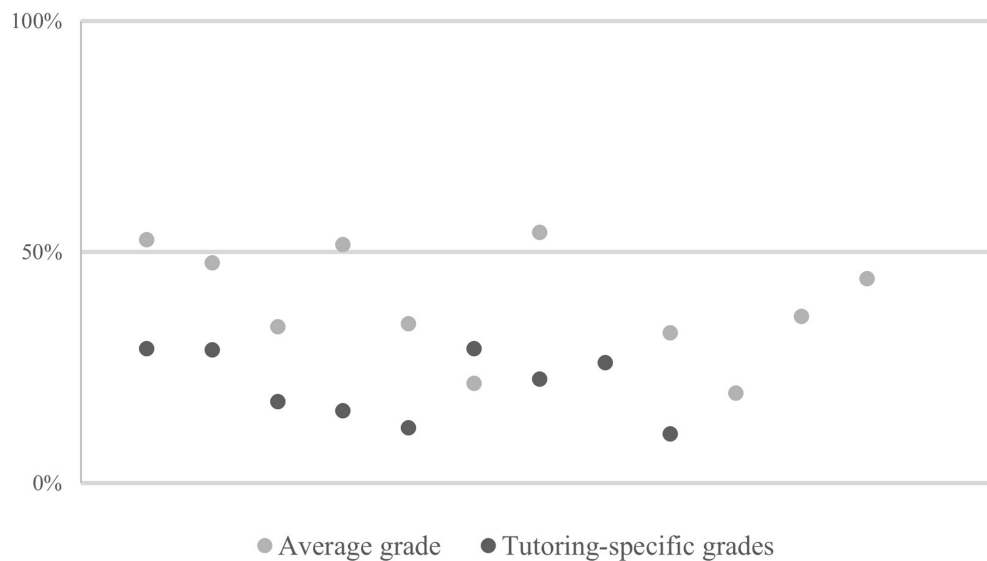


FIGURE 1
 Estimated percentage of students not participating in a catch-up program with a lower grade than students participating in a catch-up program. Each circle represents one catch-up program. The lower the percentage, the more specific the program was aimed at lower-performing students. Darker circles represent grades for the specific academic subject that was central to the catch-up program.

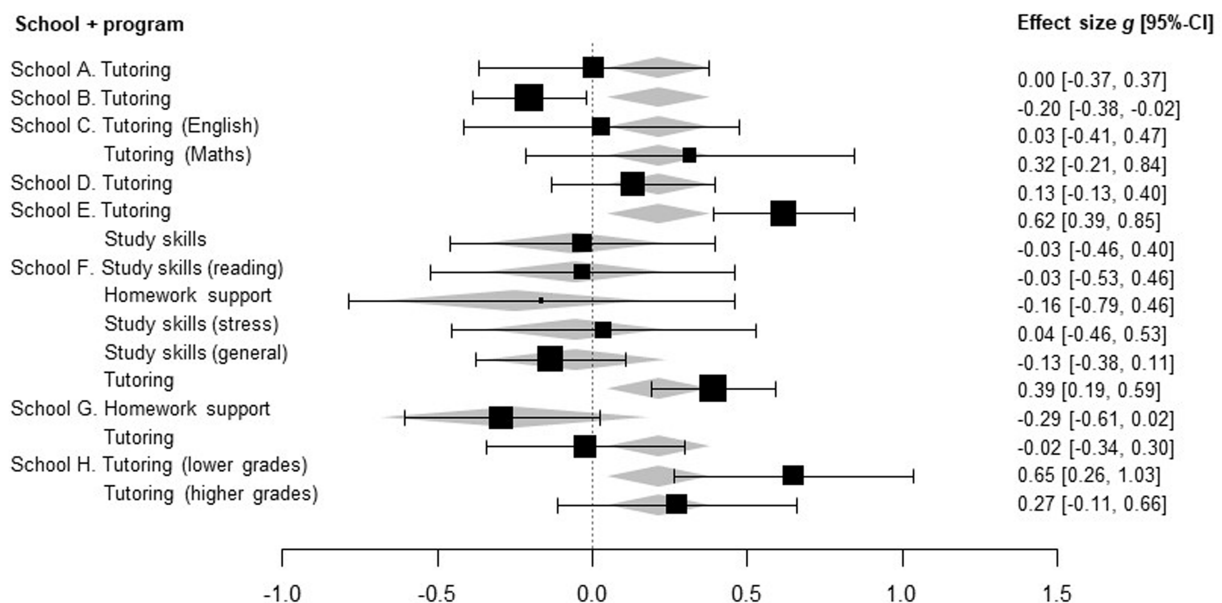
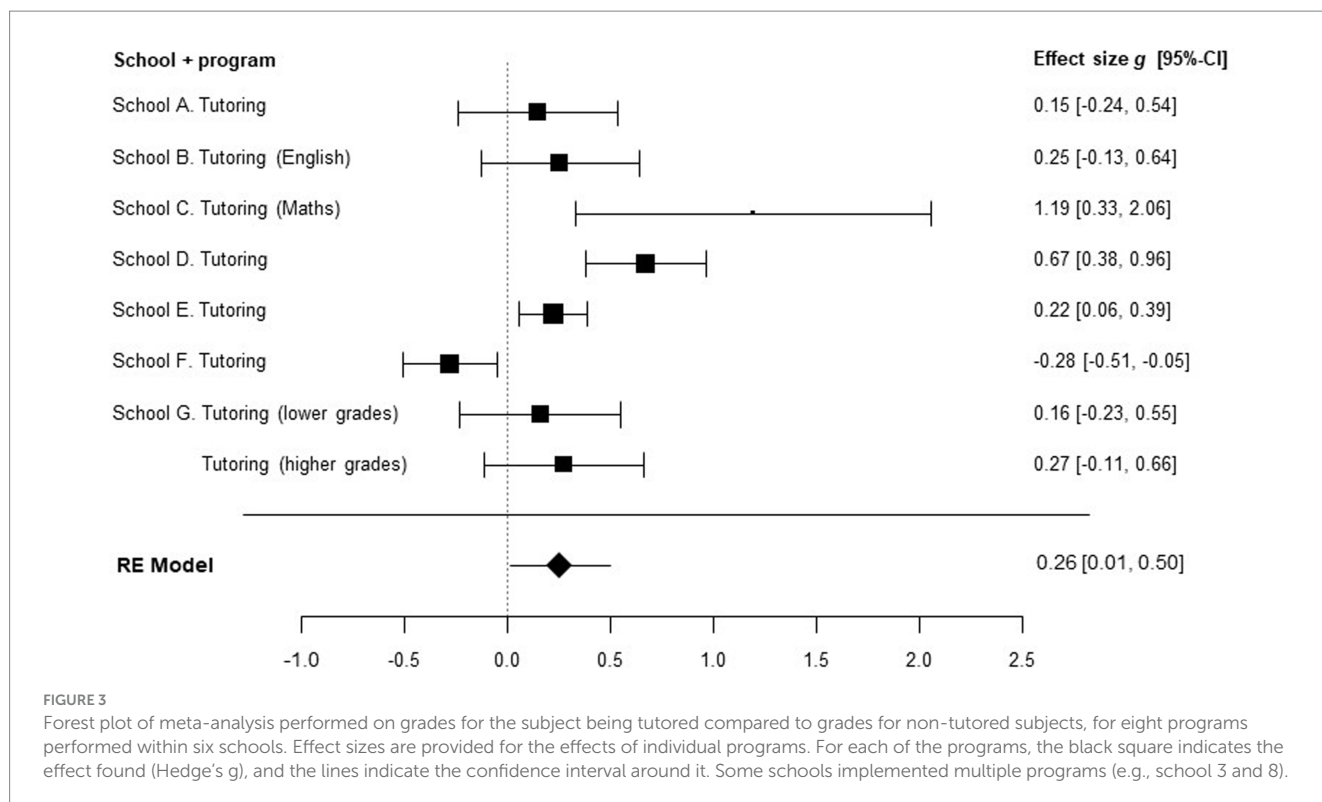


FIGURE 2
 Forest plot of meta-analysis performed on grade comparisons between participating and non-participating students in 16 intervention programs performed within eight schools. Effect sizes are provided for the effects of individual programs. For each of the programs, the black square indicates the effect found (Hedge's *g*), and the lines indicate the confidence interval around it. The gray diamond represents the estimate of the effect, given the type of intervention (tutoring – estimated effect $d = 0.26$; general skills – estimated effect $d = -0.06$; homework guidance – estimated effect $d = -0.25$). Some schools implemented multiple programs (e.g., school 3).

grades in other subjects. Importantly, there was large heterogeneity in both effect sizes across programs, as can be seen in Figures 2, 3. Although in general the programs seem to have been effective, this effectiveness differed for individual programs.

For the general skills programs and the homework support programs, effects on grades of participating students were not

significantly different from zero (see Figure 2). In fact, the point estimate of effect sizes was negative for both programs. While for both categories the confidence interval for the true effect size was large due to few effects being included, it was consistent with merely small positive effects (and large negative ones).



Discussion

This study aimed to examine the effectiveness of catch-up programs in remediating COVID-19-related learning losses among Dutch secondary school students. Firstly, we examined whether schools selected student most in need for the programs (i.e., students with the lowest grades at the start of the program). Secondly, we examined whether programs were effective in catching up learning losses, by comparing grades of participating students to those of non-participating peers; and, for subject-specific programs, by comparing students' grades in the program-specific subjects to their grades in other subjects. In doing so, we also examined differences in effectivity between different types of programs.

Participating students

Students participating in the programs in general had lower grades than their non-participating peers; and students selected for subject-specific tutoring programs had lower grades specifically in the academic subject of focus. Thus, schools seem to have selected students who were most in need of the catch-up programs, meaning that these programs reached their intended audience. This result is in line with findings in primary school, where also the lowest performing students were most likely to participate in the catch-up programs (Haelermans et al., 2021). It should be noted that we only looked at students' grades at the start of the program, meaning that schools could have just selected all students with the lowest grades, independent of whether these were the students showing the smallest learning growth during the period of school closures. Yet, previous research has shown that disadvantaged students were also the most

likely to experience adverse effects due to school closures (Engzell et al., 2021; Hammerstein et al., 2021; Haelermans et al., 2022a; Betthäuser et al., 2023; Schuurman et al., 2023), meaning that both groups possibly constitute the same students. As remediation programs seem to be most effective for the lowest performers (Education Endowment Foundation [EEF], 2021; Hammerstein et al., 2021), schools seem to have selected the right target audience for their programs.

One finding that is less in line with the idea of providing support to the weakest students is that, for half the schools, students participating in the tutoring programs only differed in their grades in the specific school subject being tutored, but not in other subjects. This finding suggests that those schools selected students weak in one subject, but not necessarily weak overall. Students may have been nominated by subject-specific teachers who might not have been aware of their students' overall GPA, but only of their subject-specific performance. Alternatively, schools may have considered it most efficient to invest in students with learning arrears in one specific subject, expecting the best results for this specific group of students.

Program effectiveness

Tutoring programs were found to have been effective in reaching their goals, as participants of these programs saw a larger increase in their grades compared to non-participants, and their grades in program-specific subjects increased more than those in non-tutored subjects. Tutoring interventions thus seemed to have helped secondary school students in catching up some of the learning losses they had experienced as result of COVID-19-related school closures.

However, general skills programs and homework support programs were not effective in increasing students' grades. While this is a null finding based on data of only few schools, the confidence interval indicated that true effect sizes ranged from substantially negative to small positive values, suggesting that strong positive effects are unlikely. The limited effectiveness of these programs for decreasing learning losses might have to do with the fact that the content of these programs was only indirectly linked to academic achievement. That is, improving grades was a distal target outcome of these programs, as the proximal outcome was to improve study skills or ability to do homework, with the expectation that this would in the end lead to better academic results as well. Possibly, students' study skills did benefit from the programs, but this did not translate into improved academic achievement, or not within a time frame of months (which was the maximum time frame examined in our study). In contrast, tutoring programs focused on achievement in one subject, which is both more directly linked to grades in that subject, and provides a more specific focus. For future studies it is advised to also measure the effect of interventions on theorized mediating constructs that might be underlying effects on academic grades.

In line with our results, recent meta-analyses have provided convincing evidence for the effectiveness of tutoring interventions in enhancing academic outcomes, reporting medium to large effect sizes (ES=0.24; Baye et al., 2019; ES=0.36; Dietrichson et al., 2017; ES=0.26; Inns et al., 2019; ES=0.37; Nickow et al., 2020; ES=0.20; Pellegrini et al., 2021). Moreover, effects of tutoring interventions seem to be larger than those of other types of programs such as computer-assisted learning or cooperative learning (Dietrichson et al., 2017; Baye et al., 2019; Inns et al., 2019; Pellegrini et al., 2021).

Yet, the effect sizes found in our study were somewhat lower than those reported in previous meta-analyses. This might be because our focus on secondary school students, whereas meta-analytic studies examined effectiveness of interventions across a range of grade levels and ages. The effects of interventions seem to be larger at the early elementary level than for higher grade levels, although few studies have examined effectiveness of tutoring programs in secondary school (Nickow et al., 2020; Education Endowment Foundation [EEF], 2021). Our effect sizes ($g=0.21$; $g=0.26$) are more in line with the 2 months learning gain (ES not reported) for tutoring program in secondary school reported in the Teaching and Learning Toolkit of the Education Endowment Foundation (EEF). Similar effects have recently been found in one of the few RCTs examining tutoring effects in secondary school (0.28SD; Guryan et al., 2023). It is generally harder to bring about learning gains in older students (Bloom et al., 2008), making smaller effect sizes very valuable, especially given that interventions other than tutoring have been found to be far less effective (Robinson and Loeb, 2021).

Differences between schools

Our results show large heterogeneity between schools, indicating that program effectiveness differed over schools, even when implementing the same type of program or having similar program goals. Program implementation thus seems to have been at least as important as the content and goals of the program. Similar types of programs may have been implemented differently, for example in duration, number of sessions, the person providing the intervention,

or enthusiasm of program personnel, all important characteristics for program effectiveness (Robinson and Loeb, 2021). For example, it has been found that programs are more effective when implemented by teachers or trained professionals compared to non-professionals (e.g., parents or volunteers), when they take place during compared to after the school day, when they have a sufficiently high dose (although it is unclear what 'sufficiently high' entails), and when they are embedded into the school curriculum (Nickow et al., 2020; Gamoran and Murnane, 2023). As there was only limited variation in these implementation measures in our study, we could not examine program characteristics determining program effectiveness in further detail. For future studies, it is advised to not only look at differences in effectivity of different broader program types, but to also examine specific characteristics of the different program types.

Also, the underlying theory about why a program works is an important determining factor for program effectiveness: the theory of change (Weiss, 1997). Programs are generally more effective with well-founded ideas about the program's target group, goal, content, and mechanisms for reaching a goal. Worryingly, most schools in our study did not have clear ideas about *how* their program would reach its goals (De Bruijn et al., 2021), making it questionable whether they followed clear guidelines on program implementation that were in line with their theory of change. Although this seems particularly problematic when using novel, non-existing programs for which no guidelines yet exist, implementation fidelity is also an important determinant of effectiveness for existing interventions (Education Endowment Foundation [EEF], 2021). To get further insight into how and which aspects of program implementation may influence program effectiveness, future studies are advised to include measures of underlying theory of change and implementation fidelity as well.

Limitations

Important strengths of this study are the meta-analytic approach, taking into account differences within and between schools, and the inclusion of a large number of secondary school students. Yet, this study also has important limitations. Firstly, despite having a large number of participating students, the number of schools in our sample was small. It can be questioned whether the sample at school level is representative of Dutch secondary schools. Also, program effectiveness might depend on certain school characteristics, such as school size or location. Given our small sample of participating schools, we were not able to take such factors into account.

Secondly, some schools implemented multiple programs simultaneously, meaning that students may have participated in more than one program at a time. Possibly, program effects are multiplicative, meaning that program effectiveness is enhanced when participating in multiple programs. Unfortunately, we were limited in inferring whether students participated in multiple programs simultaneously, meaning we only have information for their participation in the single programs. Also, again the small sample of participating schools restricted us in examining and comparing additive effects of different program combinations.

Thirdly, in the Netherlands (and elsewhere), there is no standardized monitoring system for academic achievement in secondary school, meaning that we were reliant upon averages of teacher-developed, unstandardized test outcomes. Such

unstandardized measures are seen as less reliable (Frisbie, 1988; Slavin and Madden, 2011; Dietrichson et al., 2017) as they can be biased due to, amongst others, differences in test content or testing conditions. Accordingly, larger effect sizes are generally found when comparing unstandardized to standardized test scores (e.g., De Boer et al., 2014; Cheung and Slavin, 2016; Wolf et al., 2020). Yet, by taking a meta-analytical approach, we largely controlled for between-school differences in measurement practices and outcomes. Also, this lowered reliability of teacher-developed tests is often more than compensated by averaging grades, which increases reliability and predictive validity (Meeter, 2023). Although standardized tests may produce more reliable outcomes, in themselves they are not necessarily more valid for evaluating intervention effectiveness, as validity also largely relies on the alignment between intervention goal and test outcome (Sussman and Wilson, 2019).

Lastly, we specifically focused on program effects on academic grades, whereas many schools also included goals related to non-academic outcomes, such as study skills or socio-emotional well-being (De Bruijn et al., 2021). Given the concern that has been raised regarding effects of COVID-19 on students' development in other domains than academic functioning, such as socio-emotional functioning (Racine et al., 2021; Samji et al., 2022), and the importance hereof for students' academic performance (Durlak et al., 2011; Domitrovich et al., 2017), it seems important to include these outcomes in future studies as well. Possibly, effects on other outcomes can also provide a mechanism by which programs affect students' academic performance (e.g., by participating in a program, students may feel more socially connected or become more motivated, in turn resulting in better academic performance; Durlak et al., 2011; Taylor et al., 2017). We included these outcome measures in our project as well, but unfortunately received too little response for reliable analyses.

Conclusion

Schools seem to have selected the students most in need of intervention to participate in their catch-up programs. Moreover, tutoring interventions seem to have helped secondary school students catch up some of the learning losses accrued during school closures due to COVID-19. General skills programs and homework support programs were not effective. Large heterogeneity was found between schools, implying that program implementation is at least as important for program effectiveness as the type and content of the program. Yet, many schools did not seem to have a clear idea on the mechanism underlying the desired effects. As program effectiveness is largely dependent upon the theory of change underlying the expected results (i.e., well-founded ideas about the program's target group, goal, content, and mechanisms for reaching a goal), it seems vital that schools receive help in constructing a theory of change when deciding upon to-be-implemented school-based programs. This way, it can be ensured that schools spend their money and energy wisely, helping them in reaching their desired goals. In the end, this will hopefully

References

Baye, A., Inns, A., Lake, C., and Slavin, R. E. (2019). A synthesis of quantitative research on reading programs for secondary students. *Read. Res. Q.* 54, 133–166. doi: 10.1002/rrq.229

result in them promoting their students' academic development in a positive way.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the first author upon reasonable request.

Ethics statement

The studies involving humans were approved by Vaste Commissie Wetenschap en Ethiek, Faculty of Behavioural and Human Movement Sciences, Vrije Universiteit Amsterdam. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

AB: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. MM: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, (Grant number: 40.5.20937.002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Bethäuser, B. A., Bach-Mortensen, A. M., and Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nat. Hum. Behav.* 7, 375–385. doi: 10.1038/s41562-022-01506-4

- Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *J. Res. Educ. Eff.* 1, 289–328. doi: 10.1080/19345740802400072
- Cheung, A., and Slavin, R. (2016). How methodological features affect effect sizes in education. *Educ. Res.* 45, 283–292. doi: 10.3102/0013189X16656615
- de Boer, H., Donker, A. S., and van der Werf, M. P. (2014). Effects of the attributes of educational interventions on students' academic performance: a meta-analysis. *Rev. Educ. Res.* 84, 509–545. doi: 10.3102/0034654314540006
- De Bruijn, A., Sincer, I., Turkeli, R., Meeter, M., and Ehren, M. (2021). *Catch-up programs to remediate learning losses: theory of change of primary and secondary schools in the second tranche of the COVID-19 subsidy arrangement. [Inhaal-en ondersteuningsprogramma's om leerachterstanden te remediëren: theory of change aanvragen PO en VO uit de tweede tranche van de subsidieregeling i.v.m. COVID-19]*. Amsterdam: LEARN! Research Institute.
- Di Pietro, G. (2023). The impact of COVID-19 on student achievement: evidence from a recent meta-analysis. *Educ. Res. Rev.* 39:100530. doi: 10.1016/j.edurev.2023.100530
- Dietrichson, J., Bøg, M., Filges, T., and Klint Jørgensen, A. M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: a systematic review and meta-analysis. *Rev. Educ. Res.* 87, 243–282. doi: 10.3102/0034654316687036
- Domitrovich, C. E., Durlak, J. A., Staley, K. C., and Weissberg, R. P. (2017). Social-emotional competence: an essential factor for promoting positive adjustment and reducing risk in school children. *Child Dev.* 88, 408–416. doi: 10.1111/cdev.12739
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., and Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev.* 82, 405–432. doi: 10.1111/j.1467-8624.2010.01564.x
- Education Endowment Foundation [EEF] (2021). Teaching and learning toolkit: small group tuition. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/small-group-tuition>
- Engzell, P., Frey, A., and Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2022376118. doi: 10.1073/pnas.2022376118
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Educ. Meas.* 7, 25–35. doi: 10.1111/j.1745-3992.1988.tb00422.x
- Gamoran, A., and Murnane, R. (2023). *Reducing educational inequality after the COVID-19 pandemic: what do we know, and what research do we need?* New York: William T. Grant Foundation.
- Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M., Dodge, K., et al. (2023). Not too late: improving academic outcomes among adolescents. *Am. Econ. Rev.* 113, 738–765. doi: 10.1257/aer.20210434
- Haelermans, C., Abbink, H., Baumann, S., Bijlsma, I., Havermans, W., Jacobs, M., et al. (2022b). *Learning growth reading comprehension restored in second COVID-year, spelling and mathematics not yet. [Leergeroei begrijpend lezen in tweede COVID-jaar hersteld, spelling en rekenen-wiskunde nog niet]*. NCO Factsheet No. 2022-2. Maastricht: Research Centre for Education and the Labour Market.
- Haelermans, C., Abbink, H., Baumann, S., Bijlsma, I., Havermans, W., Jacobs, M., et al. (2022c). *Less delay in learning growth of disadvantaged students in second COVID-19 year. [Minder vertraging leergeroei kwetsbare leerlingen in tweede jaar COVID-19]*. NCO Factsheet No. 2022-3. Maastricht: Research Centre for Education and the Labour Market.
- Haelermans, C., Jacobs, M., and Meeter, M. (2021). *Effects of IOP on learning growth in primary schools [Effecten van IOP op leergeroei in het PO]*. Amsterdam: LEARN! Research Institute.
- Haelermans, C., Korthals, R., Jacobs, M., de Leeuw, S., Vermeulen, S., van Vugt, L., et al. (2022a). Sharp increase in inequality in education in times of the COVID-19 pandemic. *PLoS One* 17:e0261114. doi: 10.1371/journal.pone.0261114
- Hammerstein, S., König, C., Dreisörner, T., and Frey, A. (2021). Effects of COVID-19-related school closures on student achievement—a systematic review. *Front. Psychol.* 12:746289. doi: 10.3389/fpsyg.2021.746289
- Inns, A. J., Lake, C., Pellegrini, M., and Slavin, R. (2019). *A quantitative synthesis of research on programs for struggling readers in elementary schools. Best evidence encyclopedia (BEE)*. Baltimore: Center for Research and Reform in Education.
- JASP Team (2023). JASP (Version 0.17.3) [Computer software].
- König, C., and Frey, A. (2022). The impact of COVID-19-related school closures on student achievement—a meta-analysis. *Educ. Meas.* 41, 16–22. doi: 10.1111/emip.12495
- Kortekaas-Rijlaarsdam, A. F., Ehren, M., and Meeter, M. (2020). *Catch-up programmes in primary education in the Netherlands: overview of programmes (first tranche) and a literature review of effectiveness of programmes*. Amsterdam: LEARN! Research Institute.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educ. Res.* 49, 241–253. doi: 10.3102/0013189X20912798
- Meeter, M. (2023). Predicting retention in higher education from high-stakes exams or school GPA. *Educ. Assess.* 28, 1–10. doi: 10.1080/10627197.2022.2130748
- Nickow, A., Oreopoulos, P., and Quan, V. (2020). *The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence*. National Bureau of Economic Research. Cambridge, MA.
- OECD (2018). *Education at a glance 2018: OECD indicators*. Paris: OECD Publishing.
- Pellegrini, M., Lake, C., Neitzel, A., and Slavin, R. E. (2021). Effective programs in elementary mathematics: a meta-analysis. *AERA Open* 7:233285842098621. doi: 10.1177/2332858420986211
- Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., and Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. *JAMA Pediatr.* 175, 1142–1150. doi: 10.1001/jamapediatrics.2021.2482
- Robinson, C. D., and Loeb, S. (2021). High-impact tutoring: state of the research and priorities for future learning. *Nat. Student Support Acceler.* 21, 1–53. doi: 10.26300/qf76-rj21
- Samji, H., Wu, J., Ladak, A., Vossen, C., Stewart, E., Dove, N., et al. (2022). Mental health impacts of the COVID-19 pandemic on children and youth—a systematic review. *Child Adolesc. Ment. Health.* 27, 173–189. doi: 10.1111/camh.12501
- Schuurman, T. M., Henrichs, L. F., Schuurman, N. K., Polderdijk, S., and Hornstra, L. (2023). Learning loss in vulnerable student populations after the first COVID-19 school closure in the Netherlands. *Scand. J. Educ. Res.* 67, 309–326. doi: 10.1080/00313831.2021.2006307
- Slavin, R. E., and Madden, N. (2011). Measures inherent to treatments in program effectiveness reviews. *J. Res. Educ. Eff.* 4, 370–380. doi: 10.1080/19345747.2011.558986
- Sussman, J., and Wilson, M. R. (2019). The use and validity of standardized achievement tests for evaluating new curricular interventions in mathematics and science. *Am. J. Eval.* 40, 190–213. doi: 10.1177/1098214018767313
- Taylor, R. D., Oberle, E., Durlak, J. A., and Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: a meta-analysis of follow-up effects. *Child Dev.* 88, 1156–1171. doi: 10.1111/cdev.12864
- Weiss, C. H. (1997). Theory-based evaluation: past, present, and future. *New Dir. Eval.* 1997, 41–55. doi: 10.1002/ev.1086
- Wolf, R., Morrison, J. M., Inns, A., Slavin, R. E., and Risman, K. (2020). Average effect sizes in developer-commissioned and independent evaluations. *J. Res. Educ. Eff.* 13, 428–447. doi: 10.1080/19345747.2020.1726537
- Zierer, K. (2021). Effects of pandemic-related school closures on pupils' performance and learning in selected countries: a rapid review. *Educ. Sci.* 11:252. doi: 10.3390/educsci11060252